ISSN 2618-1576 Vol. 6, Issue 3: 61-69, DOI: 10.25236/FSR.2024.060308

Research on the prediction of security fitness scale for large sports events based on machine learning algorithms

Yuanning Li[1,a], Minghui Song[1,b], Xiaojun Wang[1,c,*]

[1]Department of Physical Education, Yanshan University, Qinhuangdao, China
[a]1105793974@qq.com, [b]1060313365@qq.com, [c]137993689@qq.com
[*]Corresponding author

Abstract: *Security is an important factor for the success of the event. In the context of new quality productivity, science and technology will continuously improve the iteration rate of productivity, and machine learning algorithms are applied to the prediction of security scale of large-scale sports events to realize the dual goals of economic and safe events. In this study, three machine learning algorithms, namely, convolutional neural network, multilayer perceptual machine and decision tree algorithm, are used to predict the security scale of the seven Summer Olympic Games from 1996 to 2021 (variables such as security expenditure, total number of security personnel, and crime rate of the host country in the past five years). The results of the study show that the decision tree algorithm fits the true and predicted values of the security scale analysis better than the other two algorithms, and the weights of the model are more accurate within the margin of error. In the prediction and analysis of security scale of large-scale sports events, the application of decision tree algorithm can efficiently and accurately provide scientific and accurate theoretical basis for the pre-planning of security work, and also provide empirical reference.*

Keywords: *major sporting events; machine learing algorithms; security; fitness for scale*

1. Introduction

Productivity is the most dynamic factor driving social development and change, and is a combination of possible potential and emerging real forces.At the Second Session of the 14th National Committee of the Chinese People's Political Consultative Conference (CPPCC) in 2024, it was proposed that science and innovation should take the lead in accelerating the formation of new quality productivity[1]. The new quality productivity is the new quality state derived from the traditional productivity under the condition of informationisation and intelligent production due to the continuous innovation and breakthrough of science and technology and the continuous upgrading and development of industry, and the core meaning is to "promote the quality with the new", and drive the high-quality development with innovation. The new quality productivity will certainly provide new motive power and new mechanism for the development of sports. Under a specific task model, the analysis and prediction of existing data and the output of the corresponding procedures, which is the basic logic of artificial intelligence technology[2]. Not only AI, but also technology empowerment has been applied to various large-scale sports events in the context of new quality productivity. Large-scale sports events usually refer to national or intercontinental level or above comprehensive sports activities, the characteristics of this kind of events include huge scale, wide influence, and a large number of participants, it is because of these characteristics, large-scale sports events often become the target of terrorist attacks, so the implementation of effective security and emergency response measures to ensure the safety of participants and spectators has become an indispensable part of the traditional sports events. For example, in the case of the Summer Olympic Games, looking back at history, since the first Olympic Games was held in 1986, 32 Olympic Games have been held internationally. The scale and influence of the Olympic Games have been expanding over these 100 years, but with the increasing complexity of the global security situation, the security of the Olympic Games has become an important challenge that must be faced in every event. In retrospect, there have been a number of security incidents, particularly at the 1976 Montreal Olympics, where one of the most serious security incidents occurred[3]. Since then, Olympic organisers have realised that they must make greater efforts to ensure the safety of participating athletes and spectators, but ten years later at the 1996 Atlanta

ISSN 2618-1576 Vol. 6, Issue 3: 61-69, DOI: 10.25236/FSR.2024.060308

Olympics, terrorists launched a series of terrorist attacks, an event that also sparked major international concern about security at the Games. Today's technological level continues to improve, in the case of financial support and abundant human and material resources, how to arrange efficient and appropriate security scale has become a hot issue in all walks of life, the security scale and budget has become a major concern in the pre-planning of large-scale sports events[4].

In this study, the data of security number scale, security investment, the number of participants in the previous Summer Olympic Games, the crime rate of the host country and other indicators of the seven Summer Olympic Games from the 1996 Atlanta Olympics to the 2021 Tokyo Olympics were selected as the prediction data. By analysing the trend of the security size of the previous Summer Olympic Games, and using convolutional neural network, multi-layer perceptron neural network, and decision tree neural network to analyse the trend and pattern. This study aims to select the security model with the highest degree of fit and optimisation, which is of some significance in guiding the pre-planning of security for other large-scale events.

2. Selection of data

In this study, the data on the scale of the number of security personnel, the investment of security funds, the number of participants in the previous Summer Olympics, the crime rate of the host country and other indicators of the seven Summer Olympics from the 1996 Atlanta Olympics to the 2021 Tokyo Olympics are selected as the predictive data,as shown in Table 1. The collected data mainly come from the website of the International Olympic Committee and the website of the host country's Olympic organising committee, and part of them come from relevant news websites, and the data are collated from other scholars' relevant literature.

Table 1: Reference data.

Date and time of organisation	Countries	Number of direct entries	Number of participating countries	Expenditure on security provisions	Total security personnel	Number of offences committed in the last five years (per 100,000 people)
19 July 1996-4 August 1996	Atlanta, USA	10.7880million people	197	81.6million$	16000	8.4
15 September 2000-1 October 2000	Sydney, Australia	10.651million people	199	196.4million$	51000	1.87
13 August 2004 - 29 August 2004	Greece Athens	10.500million people	202	1791.2million$	100000	1.05
8 August 2008 - 24 August 2008	China Beijing	11.438million people	205	298.7million$	98500	1.46
27 July 2012 - 13 August 2012	London, United Kingdom	10.568million people	205	865.9million$	41700	1.11
5 August 2016 - 21 August 2016	Rio de Janeiro, Brazil	11.238million people	205	890.8million$	85000	28.6
23 July 2021-13 August 2021	Japan Tokyo	11.669million people	204	1906.1million$	78100	0.26

3. Research Methods for Machine Learning Algorithm Prediction

Neural networks have become the current mainstream in computer tasks, and models based on CNNs and MLPs are beginning to lead the way due to the promising results they have achieved in recent years, and both of these neural networks are increasingly being used for sequence prediction.

ISSN 2618-1576 Vol. 6, Issue 3: 61-69, DOI: 10.25236/FSR.2024.060308

However, decision tree models are actually playing a key role in various fields due to their higher efficiency, faster training speed and shorter training period. The three machine learning algorithms have their own characteristics. Based on this, the three machine learning algorithms selected in this study are shown in Table 2 for predicting and analysing the security scale of major sports events.

Table 2: Machine learning algorithms.

Machine Learning Algorithms	synopsis
MLP	Multilayer Perceptron (MLP) is a basic feed-forward neural network model, whose theoretical foundation is mainly based on neuron model and back propagation algorithm, including neuron model, feed-forward propagation, back propagation algorithm, etc. It achieves feature extraction and representation learning of complex data through multi-layer stacked neuron and nonlinear activation function, and thus is widely used in machine learning tasks such as classification and regression[6]. Advantages include its strong nonlinear modelling ability; applicability to a variety of tasks; ability to handle large-scale data; strong generalisation ability; and ease of implementation and debugging[5].
CNN	Convolutional Neural Network (CNN) is an artificial neural network specifically designed to process data with a grid-like structure, the core of which is to extract features from the input data through convolutional and pooling layers and to achieve tasks such as classification or regression through a fully connected layer, with the advantage of being able to take advantage of the local correlation and translational invariance of the data, which allows it to perform better than traditional neural networks when dealing with structured data.
Decision Tree	Decision tree is a machine learning algorithm based on tree structure for solving classification and regression problems. Its advantages include less data preprocessing, the ability to capture non-linear relationships between features, better results for datasets with complex interactions between features, and its automatic feature selection, which can be ranked according to the importance of the features, helping to identify the most critical features for classification or regression[6].

3.1 MLP Multi-Layer Perception Machine

Multi-Layer Perceptron, MLP (Multi-Layer Perceptron), uses a feed-forward architecture consisting of units arranged in layers. Each layer consists of nodes, and each MLP consists of at least three layers, including an input layer, one or more hidden layers, and an output layer, where the input nodes have a linear activation function and no threshold. In addition to weights, each hidden unit node and each output node has a threshold value associated with it. The hidden unit nodes have non-linear activation functions and the outputs have linear activation functions. Thus, each signal fed to a node in a subsequent layer multiplies the original input by a weight with an added threshold and then passes through a linear or nonlinear activation function (hidden unit), and the 3-layer network MLP is shown in Figure. 1.

ISSN 2618-1576 Vol. 6, Issue 3: 61-69, DOI: 10.25236/FSR.2024.060308

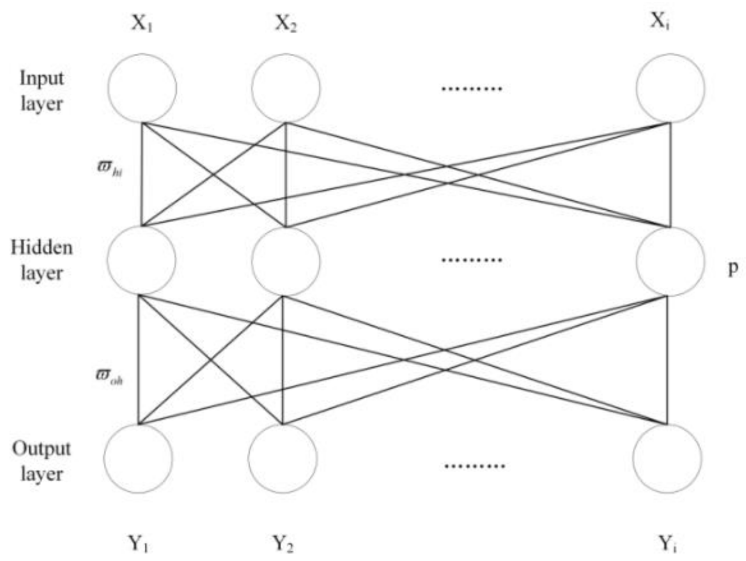

Figure 1: Schematic diagram of the MLP process.

In this paper, the input is defined as x(k), and thresholding on the hidden and output units is handled by assigning the value 1 to the augmented vector component denoted by x(N+1), and the input of the pth hidden unit is denoted as

$$T(P)=\sum_{k=1}^{N+1} w_{hi}(p,k)\cdot x(k) \tag{1}$$

The pth output representation is shown in Eq.

$$y(p)=\sum_{k=1}^{N+1} w_{op}(p,k)\cdot x(k)+\sum_{q=1}^{N_h} w_{oh}(p,q)\cdot act(q) \tag{2}$$

Where act() represents the activation function, denotes the weight of connecting the kth input unit to the pth hidden unit. denotes the weight of connecting the qth input unit to the pth hidden unit, and MSE (Mean Square Error) is used as the loss function for training.

The three-layer MLP neural network model used in this study, the input feature variables are the number of participants, the security budget, the crime rate and the number of participating countries, after the training and optimisation of the neural network, the corresponding optimal state of the weights are obtained as shown in Table 2, in which the first data -0.3408 indicates the weight from the first input unit to the first hidden unit, and so on. In this paper, the input variable is 4 and the hidden layer is 16, so the input feature matrix shape is (4, 16) and the output is (16, 1), and Relu is used as the activation function in the middle.

Table 3: Information on the weight distribution of the input samples.

Year	Number of participants	Security budget	crime rate	Number of participating countries
1996	-0.3408	-0.2229	0.1671	0.0406
2000	0.0860	-0.4068	-0.0446	0.0245
2004	0.1363	0.2593	-0.2635	-0.2541
2008	-0.0304	-0.2511	-0.0729	0.0642
2012	0.1822	-0.2897	0.2301	-0.4489
2016	-0.3794	-0.1870	0.1407	0.1375
2021	-0.2260	0.1553	0.1710	0.2074

ISSN 2618-1576 Vol. 6, Issue 3: 61-69, DOI: 10.25236/FSR.2024.060308

3.2 1D-CNN Convolutional Neural Network

Convolutional Neural Networks were initially designed to process gridded data, in this paper we use one-dimensional convolution, unlike traditional 2D-CNNs where the convolution kernel is moved in one direction to perform the convolution operation, 1D-CNN uses convolutional layers to extract features from sequential data. The convolutional layer performs a convolutional operation on the input data by sliding a fixed-size window, extracts features within the window, and then maps these features to the next layer, as shown in Figure 1. Similar to two-dimensional convolutional neural networks (2D-CNNs), 1D-CNNs can also use a pooling layer to reduce the dimensionality and computational effort of feature mapping. Our input sequence is denoted as, convolutional filter, which generates k-dimensional feature vectors through a feature window of h. The process of generating feature ci is shown in Equation (1).

$$c_i = f(\omega \cdot x_{i,i+h-1} + b) \tag{3}$$

where b represents a bias vector and f represents a nonlinear eigenfunction for filtering sequences $\{x_{1,h}, x_{2,h+1} \ldots x_{n-h+1,n}\}$ A sequence of features is formed as shown in Eq.

$$c = \{c_1, c_2, \ldots c_{n-h+1}\} \tag{4}$$

included among these $c \in R_{n-h+1}$, we then apply a maximum pooling operation and take the maximum value as the feature of this filter.

Here, we take as inputs the time of the organisation, the number of participants, the number of participating countries, the expenditure on security, and the crime rate over the last five years $x_i = \{x_1, x_2, x_3, x_4\}$ The total number of security personnel is predicted by 1D-CNN training, and the results are shown in Figure. 2.

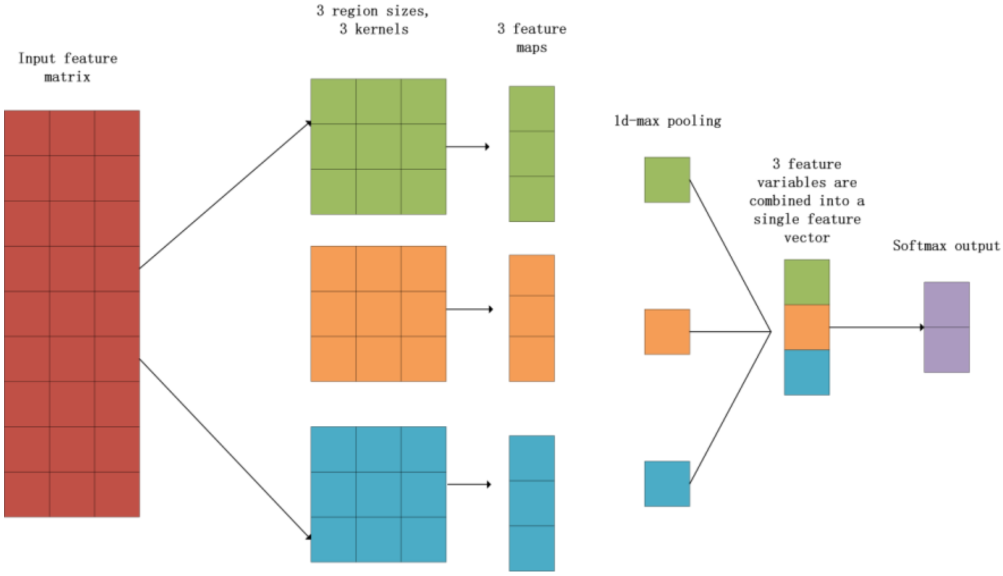

Figure 2: Schematic diagram of the 1d-conv process.

3.3 Decision Tree

Decision tree is a supervised machine learning algorithm, based on the tree structure of the model, through the recursive partition of the data, and ultimately in the leaf nodes to make predictions, as shown in Figure.3. In this paper, we use the regression decision tree model, given the input feature vector $x_i = \{x_1, x_2, x_3, x_4\}$ and define two regions $R_1(j,s) = \{x | x_j \le s\}$, $R_2(j,s) = \{x | x_j > s\}$ A heuristic method is used for the division of the feature space, where each division examines all the values of all the features in the current set one by one, and selects the optimal one of them as the cut-off

ISSN 2618-1576 Vol. 6, Issue 3: 61-69, DOI: 10.25236/FSR.2024.060308

point according to the squared error minimisation criterion, and solves for Eq.

$$\min_{j,s}\left[\min_{c_1}\sum\nolimits_{x_i\in R_1(j,s)}(y_i-c_1)^2+\min_{c_2}\sum\nolimits_{x_i\in R_2(j,s)}(y_i-c_2)^2\right] \quad (5)$$

where j denotes the cut-off variable and s denotes the cut-off point, the minimum pair of (j, s) that satisfies Eq. minimises the target division region and determines the corresponding output value, repeating the above steps for the two sub-regions until the stopping condition is satisfied, and then dividing the input space into m regions $\{R_1, R_2, ...R_m\}$ Generate a decision tree and use regression decision tree prediction to output the number of security personnel yi.

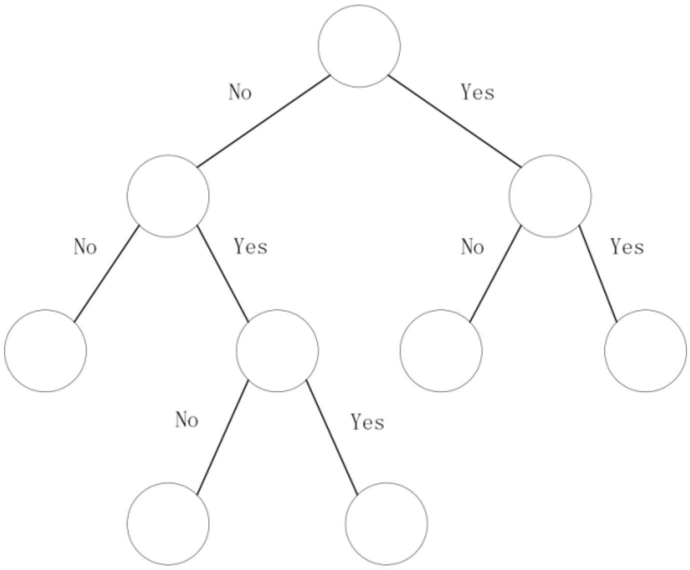

Figure 3: Schematic diagram of the decision tree.

4. Data processing

Based on the predictions of the three machine learning algorithms, the resulting predictions are shown in Table 4. The whole is shown in Figure 4 and Figure 5. The deviation of the values predicted by the three different machine learning algorithms for security size from the true value can be visualised.

Table 4: Comparison of predicted and real value data.

	1996	2000	2004	2008	2012	2016	2021
real data	16000	51000	100000	98500	41700	85000	78100
MLPPredictive data	16522	52193	108763	98972	75521	85633	78355
1D-CNNPredictive data	18224	45692	80223	97343	81544	87880	86220
Decision Tree Predictive data	1.6130	51223	100621	98511	71373	85118	78083

From the overall analysis, it can be seen that MLP is only high in the prediction of the number of security personnel in the 2004 Athens Olympic Games and the 2012 London Olympic Games, and the prediction of the rest of the years is relatively fit; 1D-CNN is not good at prediction, and Decision Tree is better, and the number of security personnel in each Olympic Games is close to the real value of the original data.

From the local analysis shown in Figure 4, we intercepted a section close to the peak of the number of security personnel to zoom in, it can be seen that the 2004 Athens Olympic Games compared with the previous two Olympic Games security personnel increased significantly, analyse the reason for this is that the 2004 Athens Olympic Games is the first Summer Olympic Games after the "9.11" incident,

ISSN 2618-1576 Vol. 6, Issue 3: 61-69, DOI: 10.25236/FSR.2024.060308

in Athens Olympic Games In the 100-day countdown to the opening of the Athens Olympic Games, there were successive explosions in front of the police station, and the security issue attracted more attention.

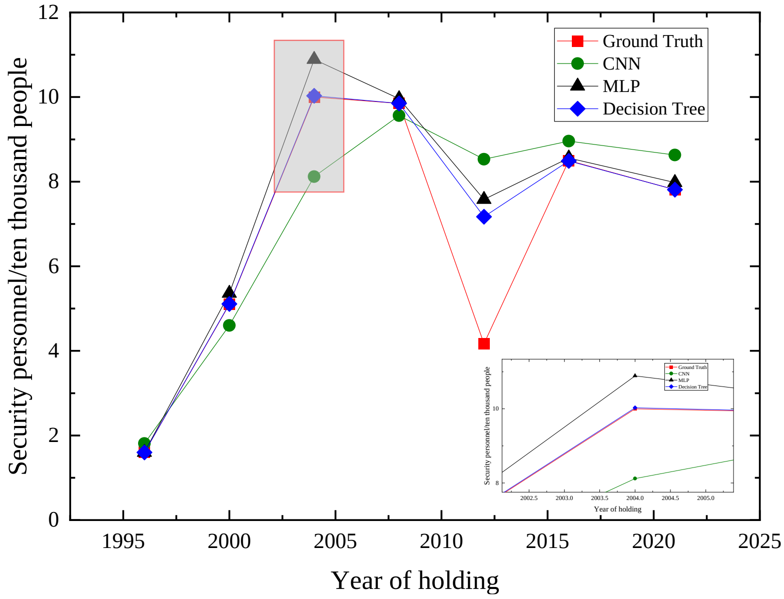

Figure 4: Security Scale Forecast (True vs. Forecast1).

From the local analysis shown in Figure 5, we intercept the three machine learning algorithms of the predicted value of the deviation from the true value of a section of the folding line to zoom in, can be seen in 2012 London Olympic Games security number of the true value of significantly lower than the predicted value of the analysis of the reasons for this is the London Olympic Games security work is not enough planning, following the "9.11" incident, the Athens Olympic Games and the Beijing Olympic Games have increased the scale of security, but the London Olympic Games security planning only 41,700 people. After the "9.11" incident, the Athens Olympic Games and the Beijing Olympic Games both increased the scale of security, but the London Olympic Games security scale planning only 41,700 people, with the opening date of the Olympic Games is gradually approaching, responsible for providing security personnel to the Olympic Games of the United States can not be promised to provide enough trained security personnel, so that the Olympic Games organisers had to turn to the military for help. In this particular case, the Decision Tree algorithm consistently provided a good fit in the prediction results.

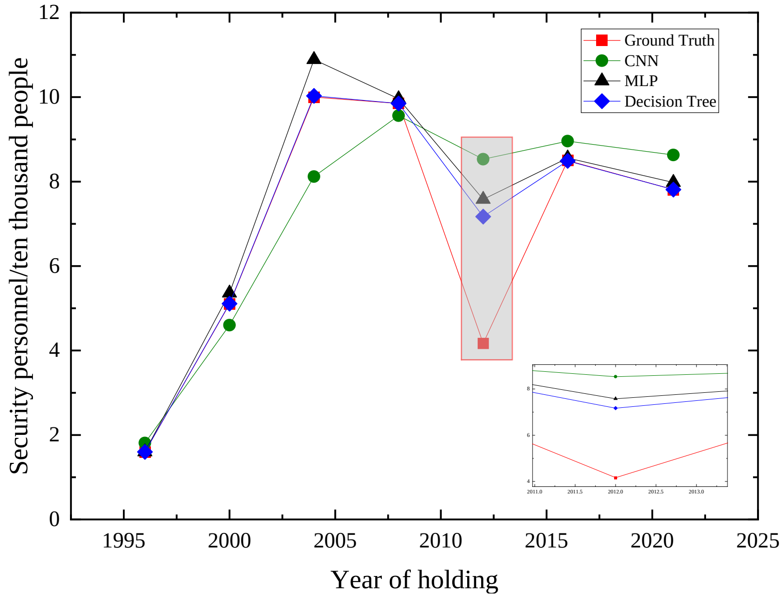

Figure 5: Security size forecast (true vs. forecast2).

ISSN 2618-1576 Vol. 6, Issue 3: 61-69, DOI: 10.25236/FSR.2024.060308

There are certain flaws in this study, the data collected is not precise enough and the limitation of related factors will produce errors in the predicted results, the lack of data volume leads to the neural network need longer time to train to ensure the model converges to the optimal solution. Accordingly, it is recommended that the management of large sports events improve the data statistics system. The data statistics system is not only to record the data of the work of the event, but more importantly, to explore its laws and trends. In all kinds of large-scale sports events, the collection and collation of security-related data, to provide a perfect data report for the retention of the security legacy, and at the same time for the scholars to do a good job in the future research on the security of large-scale sports events to provide a good basis for protection. Perfect and detailed security data reports and information can provide scientific and precise operation plans and management modes for the planning of the security scale of subsequent large-scale sports events.

5. Conclusions and Recommendations

5.1 Conclusions

In this study, three machine learning algorithms are used to predict the security scale of the previous seven Summer Olympic Games, and in the process of testing the model, the deviation of the predicted value of the decision tree algorithm for the number of security personnel from the real value is extremely small, which indicates that the algorithm's accuracy is good, and it has a practical reference value for the pre-planning of the security of large-scale sports events. Since this study only predicts the number of security personnel and does not involve the rationalisation of the allocation of security departments and other key aspects. In subsequent studies, more data from large-scale sports events can be used to predict the security scale, so that the sample size can be enlarged and more systematic predictions and analyses can be carried out on various aspects of security.

5.2 Recommendations

Precise security budget inputs. The security budget is an advance preparation for the funding and planning of security. It greatly affects the operation of security work. From the theoretical level of scientific and accurate prediction of the security budget can help the operation department to optimise the allocation of resources, improve resource utilisation. And accurate and reasonable security budget can provide decision-making and management reference basis for the operation department, to ensure that the whole security work of the systematic and scientific. At this stage, science and technology and security are closely related, the economic investment of security work has increased significantly, how to reasonably arrange the security budget has become an important link.

Optimising performance assessment. Performance assessment is the metric for managing security budgets and resources. Performance assessment of security work provides insight into how well the budget is being used and how efficiently resources are being utilised. Quantitative analysis of the performance of the security work, so as to understand the strengths and weaknesses of the work, find problems and timely adjustment and improvement, easy to compare and summarise the experience, for future budget planning and resource allocation to provide reference. Large sports events in the management of security budgets and resources, budget planning, resource allocation and performance evaluation are inseparable. Through reasonable budget planning, it can provide a clear target and reference for resource deployment; through effective resource deployment, it can achieve the maximum use of the budget; and through performance evaluation, it can find problems and shortcomings to improve efficiency.

References

[1] Xiao Feng, Shen Jianhua. Research on safety prevention of large-scale sports events[J]. Sports Culture Guide, 2007(08):52-56.
[2] Li Bin. Lessons learnt and inspiration from the Tokyo Olympic Games[J]. Hangzhou, 2022 (02): 38-43.
[3] Lv Jingjing, Song Na,Guo Qing. Construction of risk communication assessment model for large-scale sports events[J]. Journal of Wuhan Institute of Physical Education,2023,57(03):21-28.
[4] Pu Biwen, Jia Hong. Structural equation model construction and empirical research on risk assessment of large sports events[J]. China Sports Science and Technology,2018,54(02):51-58.

ISSN 2618-1576 Vol. 6, Issue 3: 61-69, DOI: 10.25236/FSR.2024.060308

[5] Su Ronghai, Niu Daoheng, Zhang Wulong. Risk assessment of safety accidents in large sports events [J]. Journal of Beijing Normal University(Natural Science Edition),2017,53(04):499-504.
[6] Zhu Dexu. Assessment and Response to Security Risks of Large-Scale Sporting Events--Taking the Security of American Sporting Events as an Example[J]. Public Security Education,2023(01):72-75.